Search result: 1 of 1

(WO/2001/001218) METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

Biblio. Data | Description | Claims | National Phase | Notices | Documents

Latest bibliographic data on file with the International Bureau

Pub. No.: WO/2001/001218 **International Application No.:** PCT/US2000/017540
Publication Date: 04.01.2001 **International Filing Date:** 26.06.2000
Chapter 2 Demand Filed: 25.01.2001

IPC: *G06F 19/00* (2006.01)

Applicants: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park New Haven, CT 06511 (US) *(All Except US).*
DENTON, Richard, Rex [US/US]; (US) *(US Only).*
JUDSON, Richard, S. [US/US]; (US) *(US Only).*
RUAÑO, Gualberto [US/US]; (US) *(US Only).*
STEPHENS, Joel, Claiborne [US/US]; (US) *(US Only).*
WINDEMUTH, Andreas, K. [DE/US]; (US) *(US Only).*
XU, Chuanbo [CN/US]; (US) *(US Only).*

Inventors: **DENTON, Richard, Rex**; (US).
JUDSON, Richard, S.; (US).
RUAÑO, Gualberto; (US).
STEPHENS, Joel, Claiborne; (US).
WINDEMUTH, Andreas, K.; (US).
XU, Chuanbo; (US).

Agent: **MOROZ, Eugene**; Morgan & Finnegan, L.L.P. 345 Park Avenue New York, NY 10154 (US).

Priority Data: 60/141,521 25.06.1999 US

Title: METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

Abstract: Methods, computer program(s) and database(s) to analyze and make use of gene haplotype information. These include methods, program, and database to find and measure the frequency of haplotypes in the general population; methods, program, and database to find correlation's between an individual's haplotypes or genotypes and a clinical outcome; methods, program, and database to predict an individual's haplotypes from the individual's genotype for a gene; and methods, program, and database to predict an individual's clinical response to a treatment based on the individual's genotype or haplotype.

Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
African Regional Intellectual Property Org. (ARIPO) (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW)
Eurasian Patent Organization (EAPO) (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM)
European Patent Office (EPO) (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE)
African Intellectual Property Organization (OAPI) (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Publication Language: English (EN)

Filing Language: English (EN)

Search result: 1 of 1

(WO/2001/001218) METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

Biblio. Data | Description | Claims | National Phase | Notices | Documents

Note: OCR Text

INTERNATIONAL SEARCH REPORT International application No. PCT/USoo/17540 C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT Category* Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. Y US 5, 773,220 A (DEKOSKY ET AL) 30 June 1998 (30-06-98), see 1-21, 30-33, 35, 43- in particular abstract and claims. 51, 53-58, 69-78,83- 84,86,94-102,104- 109,120-129, 134- 135,137,145-15 3,155-160,171-183 Y, P US 5, 972, 614 A (RUANO ET AL) 26 October 1999 (26-10-99), see 1-21, 30-33, 35, 43- in particular abstract ; claims ; column 6, lines 33-55 ; column 12, 51, 53-58, 69-78,83- lines 10-26,84,86,94-102,104- 109,120-129,134- 135, 137, 145-15 3,155-160,171-183 Y, P US 6,022,683 A (POIRIER) 08 February 2000 (08-02-00), see in 1-21, 30-33, 35, 43- particular abstract and claims. 51, 53-58, 69-78,83- 84,86,94-102,104- 109,120-129,134- 135,137,145- 153, 166-160, 171- 183 Y, P US 6,043,040 A (ACTON) 28 March 2000 (28-03-00), see in 1-21, 30-33,35,43- particular abstract, claims, and columns 49-59. 51, 53-58, 69-78,83- 84,86,94-102,104- 109,120-129,134- 135,137,145- 153, 155-160, 171- 183 Y US 5, 648,482 A (MEYER) 15 July 1997 (15-07-97), see in 1-21, 30-33,35,43- particular abstract, claims, and columns 23-26. 51, 53-58, 69- 78,8384,86,94- 102,104-109,120- 129, 134-135, 1 37,145-153,155- 160,171-183 Y, P US 6,030,778 A (ACTON ET AL) 29 February 2000 (29-02-00), see 1-21, 30-33, 35, 43- in particular abstract, claims, and columns 25-30. 51, 53-58, 69-78,83- 84,86,94-102,104- 109,120-129,134- 135,137,145- 153, 155-160, 171- INTERNATIONAL SEARCH REPORT International application No. PCT/USoo/17540 C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT Category* Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. Y gLEYN et aL Genetio Variation as a Guide to Drug Development 1-21, 30-33, 35, 43- Science. 18 September 1998. VoL 281, pages 1820-1821, see entire 51, 53-58, 69-78,83- document84,86,94-102,104-109,120-129,134- 135,137,146- 153, 166-160, 171- 183 Y MORI et aL HLA A Gene and Haplotype Frequencies in the North 1-21, 30-33, 35, 43- American Population. Transplantation. 15 October 1997, VoL 64,51, 53-58,69-78,83- No. 7, pages 1017-1027, see entire document 84,86,94-102,104- 109,120-129,134- 135,137,145- 153, 166-160, 171- 183 Y MORI et al. Computer program to predict likelihood of finding an 1-21, 30-33, 35, 43- HLA-matched donor : Methodology, validation, and application. 51, 53-58, 69-78, 83- Biology of Blood and Marrow Transplantation. October 1996, VoL 84,86,94-102,104- 2, pages 134-144, see entire document 109,120-129,134- 135,137, 145- 153, 155-160, 171- 183 Y MATISE, T. CL Genome Scanning for Complex Disease Genes 12L30-33, 36, 43- Usine the Transmission/Disequilibrium Test and Haplotype-based 51, 53-58, 69-78,83- Haplotype Relative Risk Genetic Epidemiology. 1995, VoL 12, 84,86,94-102,104- No. 6, pages 641-646, see entire document. 109,120-129,134- 135, 137, 145- 153, 155-160, 171- 183 Y COOPER et aL Network Analysis of Human Y Microsatellite 1-21, 30-33, 35, 43- Btaplotypes. Human Molecular Genetics. 1996, VoL 5. No. 11,51,53-58,69-78,83- pages 1759-1766, see entire document 84,86,94-102,104- 109,120-129, 134-135,137,145- 153, 155-160, 171- 183 Y GENE et al. Haplotype frequencies of eight Y-chromosome STR 1-21, 30-33,3543,- loci in Barcelona (North-East Spain). International Journal of 51, 53-58, 69-78,83- Legal Medicine. 1999, VoL 112, pages 403-405, see entire document 84,86,94-102,104- 109,120-129,134- 135,137,145- 153,155-160,171- 1149 INTERNATIONAL SEARCH REPORT International application No. PCT/USoo/1 7540 C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT Category* Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. Y CLARE et aL Haplotype Structure and Population Genetic 1-2t30-33, 36, 43- Inférences from Nucleotide-Sequence Variation in Human 51, 53-58, 69-78,83- Lipoprotein Lipase. American Journal of Human Genetios 1998,84,86,94-102,104- VoL 63, pages 696-912. see entire document. 109,120-129,134- 135,137,145-15 3,155-160,171-183 Y CASMIAN et aL The Irish cystic fibrosis database. Journal of 1-2130-33, 35, 43- Medical Genetics. 1995, VoL 32, No. 12, pages 972-976, see entire 61, 63-68, 69-78, 83- document.84,86,94-102,104- 109,120-129,134-135,137,145- 153, 155-160, 171- 183 Y, P TISHKOFF et aL The Accuracy of Statistical Methods for 1-21, 30-33, 35, 43- Estimation of Haplotype Frequencies : An Example from the CD4 51, 53-5869-78, 83- Locus. American Journal of Human Genetics. August 2000, VoL 84,86,94-102,104- 67, No. 2, pages 618-622, see entire document 109,120-129,134-135,137,145- 153, 155-160, 171- 183 Y PERLIN et aL Toward Fully Automate Genotyping : Allele 1-21, 30-33, 35, 43- Assignment, Pedigree Construction, Phase Determination, and 51, 53-58, 69-78,83- Recombination Detection in Duchenne Muscular Dystrophy. 84,86,94-102,104- American Journal of Human Genetics. 1994, VoL 55, No. 4, pages 109,120-129,134- 777-787, see entire document 135, 137,145- 153, 155-160, 171- 183 Y HOANG et aL PAH Mutation Analysis Consortium Database: A 1-81, 30-33, 35,43- Database for I ? isease-prodncing and Other Allelic Variation at the 51, 53-58, 69-78,83- Human PAH Locus. Nucleic Acids Research. 1996, VoL 24, No. 1,84,86,94-102,104- pages 127-131.

see entire document 109,120-129,134- 135,137,145- 153, 155-160, 171- 183 Y, P STEPHENS et aL Single-nucleotide Polymorphisms, Haplotypes, 1-21, 30-33,35,43- and Their Relevance to Pharmacogenetica. Molecular Diagnosis. 51, 53-58, 69-78,83- December 1999, VoL 4, No. 4, pages 309-317, see entire document 84,86,94-102,104- 109,120-129,134-135,137,145- 153, 155-160, 171- 183 INTERNATIONAL SEARCH REPORT International application No. PCT/USoo/17540 Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet) This international report has not been established in respect of certain claims under Article 17 (S) (a) (or the following reasons : 1.n Claims Nos because they relate to subject matter not required to be searched by this Authority. : Claims Nos. : because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically : sClaims Nos. : because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6. 4 (a Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet) This International Searching Authority found multiple inventions in this international application, as : Please See Extra Sheet. 0 As dl required additional search fees were timely paid by the applicant, this international search report covers all searchableclaims 2 As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee. s. X As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos. : 1-21, So-SS, S5, +S-51, 5S-58, 69-78,8 S-84,86,94-102,104-109,120-129,1S4-IS5, 1S7, 145-15S, 155-160,171-183 4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos. : Remark on Protest j The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees. INTERNATIONAL SEARCH REPORT International application No. PCT/USoo/17540 B. FIELDS SEARCHED Electronic data bases consulted (Name of data base and where practicable terms used) : DIALOG (files 5, 155) and EAST (files U. S. Patents, European abstracts, Japanese abstracts, and Derwent) search terms : pharmacogenomic, pharmacogenetic, haplotype, genotype, database, computer, clinical trial, population genetics, polymorphism, SNP, Hardy-Weinberg, Mendelian, linkage, phylogenetic, pedigree, locus, gene, phased, unphased BOX 11. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows: This application contins the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13. 1. In order for all inventions to be searched, the appropriate additional search fees must be paid. Group 1, claim (s) 1-8, 69-72, and 120-124, drawn to a method of generating a haplotype database, computer-usable medium, and computer programmed therefore. Group II, claim (s) 9-12 and 7s, drawn to a method of predicting the presence of a haplotype and computer-usable medium therefore. Group 111, claim (s) 1s-21, 74-78, and 125-IQ9, drawn to a method of identifying correlation between haplotype pair and clinical response, computer-usable medium, and computer programmed therefore. Group IV, claim (s) 22-29, 79-82, ISo-ISS, drawn to a method for determining susceptibility to a condition/disease, computer-usable medium, and computer programmed therefore. Group V, claim (t) 30-33, 83-84, and 134-135, drawn to a method for predicting response to treatment, computer-usable medium, and computer programmed therefore. Group VI, claim (s) 34, 86, and 136, drawn to a method for generating a tree structure, computer-usable medium, and computer programmed therefore. Group VII, claim (s) 95, 86, and 197, drawn to a method for displaying haplotype pair frequency, computer-usable medium, and computer programmed therefore. Group VIII, claim (s) S6-s7, 87-88, and 138-139, drawn to a method for displaying a linkage screen, computer-usable medium, and computer programmed therefore. Group IX, claim (s) 38-40, 89-91, and 1vu42 drawn to a method for displaying a phylogenetic tree screen, computer- usable medium, and computer programmed therefore. Group X, claim (s) 41-42, 92-9s, and 143-144, drawn to a method for displaying genotypic analysis, computer-usable medium, and computer programmed therefore. Group XI, claim (s) 43-51, 94-102, and 145-153, drawn to a method to displaying clinical response values, computer- usable medium, and computer programmed therefore. Group XII, claim (s) 52, los, and 154, drawn to a method for carrying out a genetic algorithm, computer-usable medium, and computer programmed therefore. Group XIII, claim (s) 5s, 104, and 155, drawn to a method for displaying correlations, computer-usable medium, and computer programmed therefore. Group XIV, claim (s) 54-55, 105-106, and 156-157, drawn to a method for conducting a clinical trial, computer-usable medium, and computer programmed therefore. Group XV, claim (s) 56-58, 107-109, and 158-160, drawn to a method for inferring genotype, computer-usable medium, and computer programmed therefore. Group XVI, claim (s) 59-68, 110-119, and 161-170, drawn to a method of determining polymorphic sites or subhaplotypes, computer-usable medium, and computer programmed therefore. Group XVII, claim (s) 171-175 and 18s, drawn to a data structure. Group XVIII, claim (s) 176-182, drawn to a method for storing and organizing biological information. The inventions listed as Groups I-XVIII do not relate to a single inventive concept under PCT Rule Is. 1 because, under PCT Rule 13. 2, they lack the same or corresponding special technical features for the following reasons : The special technical feature of each method is the starting materials, method steps, and goal of each method. The corresponding computer-usable medium and computer programmed therefore form part of the inventive concept with each method. Note that PCT Rule IS does not provide for multiple methods or products.

7) E. Rich and K. Knight,"Artificial Intelligence", 2"d Edition (McGraw-Hill, New York, 1991).

8) A. Ecof and B. Smouse, Genetics Vol. 136, pp. 343-359 (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within species: molecular variance parsimony.

9) G. Ruano, K. Kidd, C. Stephens, Proc. Nat. Acad. Sci., Vol. 87,6296-6300 (1990), Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules.

10) A. G. Clark, et al., Am. J. Hum. Genet., Vol. 63,595-612 (1998), Haplotype Structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art.

Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

Modifications of the above described modes for carrying out the invention that are obvious to those of skill in the fields of chemistry, medicine, computer science and related fields are intended to be within the scope of the following claims.

TABLE OF CONTENTS

CLAIMS We claim: 1. A method of generating a haplotype database for a population, comprising data elements representative of the haplotypes for at least one locus from the individuals in the population, the method comprising: (a) for each individual in the population, generating polymorphism and haplotype data elements representative of the individual's polymorphisms and haplotypes for the locus; and 1) (b) storing the polymorphism and haplotype data elements for the individuals in a computer-readable database, wherein the data elements are organized according to the spatial relationships between the polymorphisms and haplotypes and a reference nucleotide sequence for the locus.

2. The method of claim 1, wherein the locus is a gene or a gene feature and the haplotype data elements represent haplotypes and haplotype pairs for the gene or the gene feature.

3. The method of claim 2, wherein the deriving step comprises ascertaining the frequency of the haplotypes and haplotype pairs according to the Hardy- Weinberg equilibrium.

4. The method of claim 2, further comprising deriving the haplotype data elementsby: (a) determining a nucleotide sequence of the gene or the gene feature from a first chromosome and a second chromosome in each individual in the population to generate a plurality of nucleotide sequences for the population; (c) aligning the plurality of nucleotide sequences for the population; (d) identifying haplotypes from the aligned sequences; and (e) selecting two haplotypes for each individual as a haplotype pair for storage in a table in the database.

5. The method of claim 4, wherein the method further comprises validating the haplotype data.

6. The method of claim 5, wherein the validating comprises correcting an observed distribution of haplotypes or haplotype pairs for effects imposed by a limited number of individuals in the population.

7. The method of claim 6, wherein the validating also comprises analyzing compliance of the observed distribution with Mendelian inheritance principles.

8. The method of claim 1, wherein the population is selected from the group consisting of a reference population, a clinical population, a disease population, an ethnic population, a family population and a same-sex population.

9. A method of predicting the presence of a haplotype pair in an individual comprising: (a) identifying a genotype for the individual; (b) enumerating all possible haplotype pairs which are consistent with the genotype; (c) accessing a database containing reference haplotype pair frequency data to determine a probability, for each of the possible haplotype pairs, that the individual has a possible haplotype pair; and (d) analyzing the determined probabilities to predict haplotype pairs for the individual.

10. The method of claim 9, wherein the identifying step comprises determining the most predictive genotyping site or sites.

11. The method of claim 10, wherein the determining includes calculating phylogenetic and/or linkage information for the reference haplotype pairs.

12. The method of claim 10, wherein the enumerating step comprises listing the possible haplotype pairs in order of their frequency in the database.

13. A method for identifying a correlation between a haplotype pair and a clinical response to a treatment, or other phenotype, comprising: (a) accessing a database containing data on clinical responses to treatments, or other phenotypes, exhibited by a clinical population; (b) selecting a candidate locus hypothesized to be associated with the clinical response or other phenotype, the locus comprising at least two polymorphic sites; (c) providing haplotype data for each member of the clinical population, the haplotype data comprising information on a plurality of polymorphic sites

present in the candidate locus; (d) storing the haplotype data; and (e) calculating the degree of correlation between haplotype pairs and the clinical response to a treatment, or other phenotype, by statistically analyzing the haplotype and clinical response data.

14. The method of claim 13 wherein step (e) is performed last.

15. The method of claim 13 wherein step (a) is performed before any one of steps (b), (c) or (d).

16. The method of claim 13 wherein step (a) is performed after steps (b), (c) and (d).

17. The method of any one of claims 13-16, wherein the treatment comprises administration of a drug or drug candidate.

18. The method of claim 17, wherein the candidate locus is a gene or a gene feature.

19. The method of claim 18, further comprising displaying or outputting the correlation.

20. The method of claim 19, further comprising calculating the statistical significance of the correlation.

21. The method of claim 20, wherein the providing haplotype data step comprises (a) providing a genotype for the individual; (b) enumerating all possible haplotype pairs which are consistent with the genotype; (c) determining a probability for each possible haplotype pair that the individual has that possible haplotype pair, by accessing a database containing frequency data for haplotype pairs in a reference population; and (d) analyzing the determined probabilities to infer the individual's haplotype pair.

22. A method for identifying a correlation between a haplotype pair and susceptibility to a condition or disease of interest, or other phenotype of interest, comprising the steps of : (a) selecting a candidate locus hypothesized to be associated with the phenotype, condition or disease of interest, the locus comprising at least two polymorphic sites, (b) providing haplotype data for the candidate locus for each member of a population having the phenotype, condition or disease of interest ("disease haplotype data"); (c) organizing the disease haplotype data in a database; (d) statistically analyzing the disease haplotype data to calculate haplotype pair frequencies; (e) accessing a database containing haplotype data for the candidate locus for each member of a healthy reference population ("reference haplotype data"); (f) statistically analyzing the reference haplotype data to calculate haplotype pair frequencies; and (g) when a haplotype pair has a higher frequency in the population having the phenotype, condition or disease of interest than in the healthy reference population, identifying a correlation of the haplotype pair with susceptibility to the disease or condition of interest.

23. The method of claim 22 wherein step (f) is performed after step (d).

24. The method of claim 22 wherein step (e) is performed before any one of steps (b), (c), or (d).

25. The method of claim 22 wherein step (e) is performed after any one of steps (b), (c), or (d).

26. The method of any one of claims 22-25, wherein the candidate locus is a gene or a gene feature.

27. The method of claim 26, further comprising displaying or outputting the identified correlation.

28. The method of claim 27, further comprising calculating the statistical significance of the identified correlation.

29. The method of claim 28, wherein the providing haplotype data step comprises: (a) providing a genotype for the individual; (b) enumerating all possible haplotype pairs which are consistent with the genotype; (c) for each possible haplotype pair, determining the probability that the individual has that haplotype pair, by accessing a database containing frequency data for haplotype pairs in a reference population; and (d) inferring the individual's haplotype pair based on the determined probabilities.

30. A method of predicting an individual's response to a medical or pharmaceutical treatment, comprising: (a) selecting at least one candidate gene for which a correlation between haplotype content and response to the treatment has been identified; (b) determining the haplotype pair of the individual for the candidate gene or genes; and (c) predicting that the individual's response will be the response associated haplotype pair with information on the correlation.

31. The method of claim 30, wherein the selecting step comprises outputting a list of candidate genes associated with different responses to the treatment.

32. The method of claim 31, further comprising storing the haplotype pair.

33. The method of claim 32, further including generating an error estimate.

34. A computer implemented method for generating a gene structure screen for display on a display device, comprising the steps of : (a) retrieving from a database and displaying in a first area data indicative of the frequencies of occurrence of a gene's haplotypes within predetermined member groupings of a reference population; (b) retrieving from a database and displaying in a second area data indicative of the frequencies of occurrence of particular nucleotides for the member groupings; (c) retrieving from a database data indicative of gene structure; (d) displaying in a third area a graphical representation of gene structure that identifies polymorphic sites on the gene; (e) selecting one of the polymorphic sites to cause the appropriate nucleotide frequencies to be displayed in the second area.

35. A computer implemented method for generating a haplotype pair frequency screen for display on a display device, comprising the steps of : (a) displaying in a first area a plurality of selectable items each corresponding to a polymorphic site for a predetermined gene; (b) selecting one or more of said selectable items; (c) displaying in a second area the haplotype pairs occurring in a reference population for the selected polymorphic sites; (d) displaying in a third area data indicative of haplotype frequencies for a plurality of member groupings within the population.

36. A computer implemented method for generating a linkage screen for display on a display device, comprising the steps of : (a) displaying in a first area a graphical scale showing a reference for determining progressive degrees of linkage between polymorphic sites in a population; (b) displaying in a second area a graphical matrix structure having a plurality of grids, where each axis of the structure represents polymorphic sites on a gene; and where each grid graphically displays an indication of degree of linkage between polymorphic sites corresponding to that grid, in accordance with the reference shown in the first area.

37. The method of claim 36, wherein color is used as the indication of degree of linkage.

38. A computer implemented method for generating a phylogenetic tree screen for display on a display device, comprising the steps of : (a) displaying in a first area a plurality of selectable items each corresponding to a polymorphic site for a predetermined gene; (b) selecting one or more of said selectable items; (c) displaying in a second area a phylogenetic tree structure having nodes for each haplotype in a population, where the distance between nodes is indicative of the number of nucleotides that would have to be flipped to change one haplotype into another.

39. The method of claim 38, wherein the nodes are connected by links that indicate a single nucleotide difference between nodes.

40. The method of claim 39, wherein the nodes each display an indication of ethnogeographic frequency of occurrence of the haplotype represented by the node.

41. A computer implemented method for generating a genotype analysis screen for display on a display device, comprising the steps of : (a) displaying a first plurality of selectable items each corresponding to a polymorphic site, and a plurality of second selectable items each corresponding to a polymorphic site; (b) displaying a graphical scale showing a reference for determining progressive degrees of haplotype identification reliability using genotyping; (c) displaying a graphical matrix structure having a plurality of grids, where each axis represents a haplotype indicated by the first selectable items; and where each grid graphically displays an indication of degree of identification reliability for identifying the haplotype corresponding to that grid using genotyping specified by the second selectable items, in accordance with the reference.

42. The method of claim 41, wherein the indication of degree is color.

43. A method of displaying clinical response values of a subject population as a function of haplotype pairs of the individuals in the population, comprising: (a) receiving from a computer-readable storage device, data representing haplotype pairs and clinical response values for the subject population; (b) graphically displaying a haplotype pair matrix each of whose cells contains a graphical representation of the clinical response values of individuals having the haplotype pair corresponding to that cell of the haplotype pair matrix.

44. A method of displaying clinical response values of a subject population as a function of haplotype pairs of the individuals in the population, comprising: (a) displaying one or more first selectable items representing polymorphic sites for a predetermined gene, which when selected, will generate haplotype pairs; (b) displaying a second selectable item representing a clinical response measurement; which, when selected in conjunction with the first selectable items will cause display of a haplotype pair matrix, each of whose cells contains a graphical representation of the clinical response values for the selected clinical measurement of individuals having the haplotype pair corresponding to that cell of the haplotype pair matrix.

45. The method of claim 43 or 44, wherein the graphical representation of clinical response values is a color scale or gray scale, the shade of each cell being proportional to the mean clinical response value of individuals having the haplotype pair corresponding to that cell of the haplotype pair matrix.

46. The method of claim 45, further comprising displaying a means for adjusting the range of mean clinical response values represented by the color scale or gray scale, wherein adjustment of the range causes the displayed shade of color or gray of the cells of the haplotype pair matrix to be adjusted accordingly.

47. The method of claim 43 or 44 wherein the graphical representation of data is a histogram indicating the distribution of individuals across the range of clinical response values.

48. The method of any one of claims 43,44, or 45 wherein at least one cell includes a selectable area which, when selected, will cause the display of a histogram indicating the distribution of individuals across the range of clinical response values.

49. The method of any one of claims 43,44 or 45 which further comprises displaying a selectable item which, when selected, causes the display of the statistical significance of the correlations between variation at individual polymorphic sites and the clinical response values.

50. The method of claim 43,44 or 45 which further comprises displaying a selectable item which, when selected, displays the numerical mean and standard deviation of clinical response values among individuals having each haplotype pair in the matrix.

51. The method of claim 43,44 or 45 which further comprises displaying a selectable item which, when selected, causes the display of the results of an analysis of variation calculation to permit determination of whether variation in the clinical response values between individuals having different haplotype pairs is statistically significant.

52. A computer-implemented method for carrying out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data to a clinical response measurement comprising: (a) displaying a variable controller for setting the number of genetic algorithm generations parameter; (b) displaying a variable controller for setting the number of agents parameter; (c) displaying a variable controller for setting the mutation rate parameter; (d) displaying a variable controller for setting the crossover rate parameter; (e) displaying one or more selectable items each corresponding to a polymorphic site of a predetermined gene; and (f) displaying a selectable item for initiation of the genetic algorithm calculation; wherein selection of one or more selectable items corresponding to a polymorphic site, and selection of the item for initiation of the genetic algorithm calculation, results in the execution of the genetic algorithm calculation with the parameters set by the variable controllers, and the display of the residual error of the model as a function of the number of genetic algorithm generations and a display of the results of the genetic algorithm calculation showing the optimal weights for each of the polymorphic sites.

53. A computer-implemented method for displaying correlations between clinical outcome values for a selected population, comprising: 2) (a) displaying a first plurality of selectable items corresponding to the clinical outcome variables; 3) (b) displaying a second plurality of selectable items corresponding to the clinical outcome variables; and 4) (c) displaying a scatter plot of data points corresponding to the individuals in the selected population; 5) wherein selecting first item from the first plurality of selectable items causes each data point to be plotted on the x axis of the scatter plot according to the value of the corresponding clinical outcome value for the individual associated with the data point, and wherein selection of a second item from the second plurality of selectable items causes each data point to be plotted on the y axis of the scatter plot according to the value of the corresponding clinical outcome value for the individual associated with the data point.

54. A method for conducting a clinical trial of a treatment protocol for a medical condition of interest, comprising: (a) selecting one or more genes (or other loci) known or expected to be involved in a particular disease or drug response; (b) defining a reference population of healthy individuals with a broad and representative genetic background; (c) sequencing DNA from each member of the reference population; (d) determining the haplotypes for each of the selected genes (or other loci) for each member of the reference population; (e) determining the frequencies, population distributions and statistical measures, including confidence limits, for each of the determined haplotypes; (f) recruiting a trial population of individuals who have the medical condition of interest; (g) treating individuals in the trial population according to the treatment protocol, and measuring their response to treatment; (h) determining the haplotypes for each of the selected genes (or other loci) for each member of the trial population; (i) determining the correlations between individual responses to the treatment and individual haplotype content for each of the selected genes (or other loci); and (j) from these correlations, constructing a model that predicts the response of an individual to the treatment, given the individual's haplotype content.

55. The method of claim 54, further comprising the step of deriving from the haplotype distribution found for the reference population a reduced set of genotyping markers, which allow an individual's haplotypes to be accurately predicted without conducting a complete molecular haplotype analysis, and using the reduced set of genotype markers to determine haplotypes in step (h).

56. A method of inferring genotypes of individual subjects for a selected gene having at least m polymorphic sites, comprising (a) providing a database of m-site haplotypes of the selected gene from a representative cohort of individuals; (b) tabulating the frequency of occurrence for each of the haplotypes; (c) constructing a list of all genotypes that could result from all possible pairs of observed haplotypes; (d) calculating the expected frequency of these genotypes assuming the Hardy-Weinberg equilibrium; (e) generating a complete set of all possible masks of the same length m as the haplotypes, wherein each mask blocks the identity of the nucleotides at m-n polymorphic sites and admits the identity of nucleotides at the other n sites; (f) for each mask, calculating how much ambiguity results from genotyping with only the n polymorphic sites whose identity is admitted by the mask; (g) from among those masks having an acceptable level of ambiguity, selecting a mask which has the lowest value of n; (h) genotyping the subjects by measuring only the n polymorphic sites that are admitted by the selected mask; and (i) assigning to each subject having a particular n-site haplotype, the full m-site haplotype of a member of the initial cohort having the same n-site haplotype.

57. The method of claim 56, wherein the calculation of ambiguity for a mask comprises (a) identifying all pairs of genotypes that are rendered identical by application of the mask; (b) calculating the geometric mean of the calculated Hardy- Weinberg frequencies of each pair of genotypes identified in step (a); (c) summing all such geometric means for all ambiguous pairs to obtain an ambiguity score for the mask.

58. The method of either of claims 56 or 57, wherein, if application of the selected screen causes an ambiguity in that two haplotype pairs A and B exist that could explain a given genotype, and the Hardy-Weinberg equilibrium predicts probabilities PA and PB, where PA+PB=1, the assignment of a haplotype pair is carried out by a process comprising (a) selecting a random number between 0 and 1; (b) if the random number is less than or equal to PA, assigning the haplotype pair A; and (c) if the number is greater than PA, assigning the haplotype pair B.

59. A method of determining polymorphic sites or sub-haplotypes that correlate with a clinical response or outcome of interest, comprising: (a) providing haplotype information, and clinical response or outcome data (clinical outcome values) from a cohort of subjects; (b) statistically analyzing each individual SNP in the haplotype for the degree to which it correlates with the clinical outcome values, and generating a numerical measure of the degree of correlation; (c) saving for further processing those individual SNPs whose numerical measure of the degree of correlation with the clinical outcome values exceeds a first cut-off value; (d) generating all possible pair-wise combinations of the saved SNPs so as to provide

a set of n-site sub-haplotypes where n = 2; (e) statistically analyzing each newly generated n-site sub- haplotype for the degree to which it correlates with the clinical outcome values and calculating a numerical measure of the degree of correlation; (f) saving for further processing those n-site sub-haplotypes whose numerical measure of the degree of correlation with the clinical outcome values exceeds the first cut-off value; (g) generating all possible pair-wise combinations among and between the saved SNPs and saved sub-haplotypes, to produce new subhaplotypes with increased values of n; (h) repeating steps (e) through (g) until either (i) no new sub- haplotypes can be generated, or (ii) no further sub-haplotypes having n less than a pre-selected limit can be generated.

60. The method of claim 59, further comprising the step of displaying those saved SNPs and sub-haplotypes whose numerical measure of the degree of correlation with the clinical outcome value exceeds a second cut-off value, wherein the second cut-off value is greater than the first cut-off value.

61. The method of claim 59, wherein the numerical measure of degree of correlation is replaced by the p-value for the correlation, and SNPs and sub- haplotypes are saved if the p-value is less than a first cut-off value.

62. The method of claim 61, further comprising the step of displaying those saved SNPs and sub-haplotypes whose p-value for the correlation with the clinical outcome value is less than a second cut-off value, wherein the second cut-off value is less than the first selected value.

63. The method of any one of claims 59-62, further comprising the step of excluding from further processing complex subhaplotypes which are constructed from smaller sub-haplotypes, where the smaller sub-haplotypes each have correlation values that are at least as significant as that of the complex sub- haplotype.

64. A method of determining polymorphic sites or sub-haplotypes that correlate with a clinical response or outcome of interest, comprising: (a) providing single gene haplotype information for one or more genes, and clinical response or outcome data, from a cohort of subjects; (b) statistically analyzing each single gene haplotype for the degree to which it correlates with the clinical response or outcome of interest, and calculating a numerical measure of the degree of correlation; (c) saving for further processing those haplotypes whose numerical measure of the degree of correlation with the clinical response or outcome of interest exceeds a first selected value; (d) for each haplotype composed of m polymorphic sites, generating all possible sub-haplotypes having a single site masked, so as to provide a set of sub-haplotypes having (m-n) sites, where n = 1; (e) statistically analyzing each newly generated sub-haplotype for the degree to which it correlates with the clinical response or outcome of interest, and calculating a numerical measure of the degree of correlation; (f) saving for further processing those sub-haplotypes whose numerical measure of the degree of correlation with the clinical response or outcome of interest exceeds the first selected value; (g) from the saved sub-haplotypes, generating all possible sub- haplotypes having one additional site masked; (h) repeating steps (e) through (g) until either (i) no new sub- haplotypes have a degree of correlation which exceeds the first selected value, or (ii) no further sub-haplotypes having more unmasked sites than a pre-selected limit can be generated.

65. The method of claim 64, further comprising the step of displaying those saved sub-haplotypes whose numerical measure of the degree of correlation with the clinical response or outcome of interest exceeds a second selected value, wherein the second selected value is greater than the first selected value.

66. The method of claim 64, wherein the numerical measure of degree of correlation is replaced by the p-value for the correlation, and sub-haplotypes are saved if the p-value is less than a fi3st selected value.

67. The method of claim 66, further comprising the step of displaying those saved sub-haplotypes whose p-value for the correlation with the clinical response or outcome of interest is less than a second selected value, wherein the second selected value is less than the first selected value.

68. The method of any one of claims 64-67, further comprising the step of excluding from further processing complex subhaplotypes which are constructed from smaller sub-haplotypes, where each of the smaller sub-haplotypes has correlation values that are at least as significant as that of the complex sub- haplotype.

69. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to adjust observed haplotype pair frequencies within a population group, said haplotype pair frequencies being stored in a

computer-readable database of haplotype information for a gene or gene feature of interest, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access said database and generate all possible haplotype pairs consistent with the stored genotypes; (b) computer-readable program code for causing a computer to calculate the expected frequency of the generated haplotypes and haplotype pairs according to the Hardy-Weinberg equilibrium, based upon the observed distribution of haplotypes or haplotype pairs in the population; and (c) computer-readable program code for causing a computer to select the most probable haplotype pair for the individual based on the observed.

70. The computer-usable medium of claim 69, further comprising computer- readable program code stored thereon for causing a computer to correct the stored distribution of haplotypes or haplotype pairs for effects imposed by the presence of a limited number of individuals in the population.

71. The computer-usable medium of claim 69, further comprising computer- readable program code stored thereon for causing a computer to validate haplotype pair assignments by analyzing for compliance of the assigned haplotype pair with Mendelian inheritance principles.

72. The computer-usable medium of claim 69, wherein the population is selected from the group consisting of a reference population, a clinical population, a disease population, an ethnic population, a family population and a same-sex population.

73. A computer-usable medium having computer-readable program code stored thereon, for causing haplotype pair assignments to be made to an individual member of a population whose genotype information for a gene or gene feature of interest is stored in a computer-readable form, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to generate all possible haplotype pairs consistent with the stored genotype; (b) computer-readable program code for causing a computer to access a database containing reference haplotype pair frequency data and to determine from the frequency data the probability, for each of the possible haplotype pairs, that the individual has the possible haplotype pair; and (c) computer-readable program code for causing a computer to select the most probable haplotype pair for the individual.

74. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to identify a correlation between a clinical response to a treatment or other phenotype and a haplotype or haplotype pair present at a candidate locus hypothesized to be associated with the clinical response other phenotype, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database containing data on clinical responses to treatments, or other phenotypes, exhibited by individuals in a clinical population; (b) computer-readable program code for causing a computer to access a database containing haplotype data for each individual of the clinical population, the haplotype data comprising information on a plurality of polymorphic sites present at the candidate locus; and (c) computer-readable program code for causing a computer to calculate the degree of correlation between haplotype pairs and the clinical response to the treatment or other phenotype, by statistical analysis of the haplotype and clinical response data.

75. The computer-usable medium of claim 74, wherein the treatment comprises administration of a drug or drug candidate.

76. The computer-usable medium of claim 74, wherein the candidate locus is a gene or a gene feature.

77. The computer-usable medium of claim 74, further comprising computer- readable program code stored thereon for causing a computer to store, display, or output the degree of correlation.

78. The computer-usable medium of claim 74, further comprising computer- readable program code stored thereon for causing a computer to calculate the statistical significance of the correlation.

79. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to identify a correlation between an individual's susceptibility to a condition or disease of interest, or other phenotype, and a haplotype or haplotype pair present at a candidate locus hypothesized to be associated with susceptibility to the condition or disease of interest, or with a phenotype of interest, the computer-readable program code comprising: (a) computer-

readable program code for causing a computer to access haplotype data for the candidate locus for each member of a population having the phenotype or condition or disease of interest ("disease haplotype data"); (b) computer-readable program code for causing a computer to statistically analyze the disease haplotype data to calculate haplotype or haplotype pair frequencies; (c) computer-readable program code for causing a computer to access a database containing haplotype data for the candidate locus for each member of a healthy reference population ("reference haplotype data"); (d) computer-readable program code for causing a computer to statistically analyze the reference haplotype data to calculate haplotype or haplotype pair frequencies; and (e) computer-readable program code for causing a computer to identify a correlation of a haplotype or haplotype pair with susceptibility to the disease or condition of interest, or with the phenotype of interest, when the haplotype or haplotype pair has a higher frequency in the population having the phenotype, condition or disease of interest than in the reference population.

80. The computer-usable medium of claim 79, wherein the candidate locus is a gene or a gene feature.

81. The computer-usable medium of claim 79, further comprising computer- readable program code stored thereon for causing a computer to store, display, or output the identified correlation.

82. The computer-usable medium of claim 79, further comprising computer- readable program code stored thereon for causing a computer to calculate the statistical significance of the correlation.

83. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to predict an individual's response to a medical or pharmaceutical treatment based on one or more selected haplotypes or haplotype pairs of the individual, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database of correlations between haplotypes or haplotype pairs and responses to the medical or pharmaceutical treatment in a reference population; (b) computer-readable program code for causing a computer to locate haplotypes or haplotype pairs in the database that match the selected haplotype pairs of the individual, and (c) computer-readable program code for causing a computer to predict that the individual's response will be the response or responses associated in the database with the selected haplotype or haplotype pair.

84. The computer-usable medium of claim 83, further comprising computer- readable program code stored thereon for causing a computer to generate an error estimate for the prediction.

85. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display a gene's structure and gene features on a display device, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to retrieve from a database, and display in a first area of the display device, data indicative of the frequencies of occurrence of a gene's haplotypes within predetermined member groupings of a reference population; (b) computer-readable program code for causing a computer to retrieve from a database data indicative of the gene's structure and gene features; (c) computer-readable program code for causing a computer to display in a second area of the display device a graphical representation of the gene's structure, user-selectable items indicating the location of gene features, and graphical indicators of the location of polymorphic sites on the gene; (d) computer-readable program code for causing a computer to display in a third area of the display device, in response to a user's selection of an item indicating a gene feature, a graphical representation of the structure of the gene feature having user- selectable items indicating the position of polymorphic sites; and (e) computer-readable program code for causing a computer to retrieve from a database, and display in a third area of the display device, in response to a user's selection of an item indicating the position of a polymorphic site, data indicative of the frequencies within the member groupings of the occurrence of particular nucleotides at the polymorphic site.

86. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display on a display device haplotype pair frequency data within a population of individuals, for a selected gene or gene feature, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to display on the display device a plurality of selectable items, each item corresponding to a polymorphic site in the gene or gene feature; (c) computer-readable program code for causing a computer to retrieve from a database and display on the display device, in response to a user's selection of one or more items indicating polymorphic sites, individual haplotype pairs in the database that differ at one or more of the selected polymorphic sites; and (d) computer-readable program code for causing a computer to display on the display device data indicative of the frequencies of the displayed haplotype pairs within one or more member groupings within the population.

87. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display on a display device polymorphic site linkage data for a gene or gene structure of interest, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to display on the display device one or more matrix structures, wherein the axes of each matrix structure represent the polymorphic sites in the gene or gene feature of interest, and wherein each matrix structure corresponds to a different population or population group; and (b) computer-readable program code for causing a computer to display on the display device, in each cell of a matrix structure, a graphical indication of degree of linkage between the twp polymorphic sites corresponding to the coordinates of the cell in the matrix.

88. The computer-usable medium of claim 87, wherein color is used as the graphical indication of degree of linkage, and wherein the medium further comprises computer-readable program code stored thereon for causing a computer to display a reference color scale relating color to degree of linkage.

89. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display on a display device a phylogenetic tree, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to display a plurality of selectable items, each corresponding to a polymorphic site in the gene or gene feature of interest; and (b) computer-readable program code for causing a computer to display a phylogenetic tree structure having a node for each haplotype in a population, where the distance between nodes is proportional to the minimum number of nucleotides that would have to be changed to interconvert the corresponding haplotypes.

90. The computer-usable medium of claim 89, further comprising computer- readable program code stored thereon for causing a computer to display connections between the nodes that indicate a single nucleotide difference between the haplotypes repesented by the nodes.

91. The computer-usable medium of claim 89, further comprising computer- readable program code stored thereon for causing a computer to display at each node an indication of the relative frequency of occurrence of the haplotype represented by the node among different population groups.

92. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display a genotype analysis screen on a display device, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to display a first plurality of selectable items, each corresponding to a polymorphic site, and a second plurality of selectable items, each corresponding to a polymorphic site; (b) computer-readable program code for causing a computer to display on the display device a matrix structure, wherein the axes of the matrix structure represent haplotypes in the gene or gene feature of interest that vary at the polymorphic sites selected from the first plurality of selectable items; and (c) computer-readable program code for causing a computer to display on the display device, in each cell of the matrix structure, a graphical indication of the reliability of the assignment to an individual of the haplotype pair corresponding to the coordinates of the cell in the matrix, when the individual is genotyped only at the polymorphic sites selected from the second plurality of selectable items.

93. The computer-usable medium of claim 92, wherein color is used as the graphical indication of reliability of haplotype pair assignment, and wherein the medium further comprises computer-readable program code stored thereon for causing a computer to display a reference color scale relating color to reliability of haplotype pair assignment.

94. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display clinical response values, or other phenotype data, of a subject population as a function of haplotype pairs of the individuals in the population, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to retrieve from a computer-readable storage device, data representing haplotype pairs and clinical response values, or other phenotype data, for the subject population; and (b) computer-readable program code for causing a computer to graphically display a haplotype pair matrix structure, each of whose cells contains a graphical representation of the clinical response values or other phenotype data of individuals having the haplotype pair corresponding to the coordinates of that cell in the haplotype pair matrix.

95. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display on a display device clinical response values, or other phnotypic data, of a subject population as a function of the haplotype pairs of the individuals in the population for a gene or gene feature of interest, the computer-readable program

code comprising: (a) computer-readable program code for causing a computer to display one or more first selectable items representing polymorphic sites of the gene of gene feature; (b) computer-readable program code for causing a computer to display one or more second selectable items representing clinical measurements or phenotypes: and (c) computer-readable program code for causing a computer to display on the display device, in response to the selection by the user of at least one first and second selectable items, a haplotype pair matrix structure, wherein the axes of the matrix structure represent haplotypes in the gene or gene feature of interest that vary at the polymorphic sites corresponding to the first selected item or items, and wherein each of the cells of the matrix contains a graphical representation of the mean clinical response value, or other phenotype data, for the clinical measurement represented by the selected second item, of individuals having the haplotype pair corresponding to the coordinates of the cell in the haplotype pair matrix.

96. The computer-usable medium of claim 94 or 95, wherein color is used as the graphical indication of mean clinical response value, or other phenotype data, and wherein the medium further comprises computer-readable program code stored thereon for causing a computer to display a reference color scale relating color to mean clinical response value.

97. The computer-usable medium of claim 96, wherein the medium further comprises: (a) computer-readable program code stored thereon for causing a computer to display a means for adjusting the range of mean clinical response values or other phenotype data represented by the reference color scale; and (b) computer-readable program code stored thereon for causing a computer, in response to the adjustment of the range of clinical response values or other phenotype data represented by the reference color scale, to adjust the color of the cells of the haplotype pair matrix.

98. The computer-usable medium of claim 94 or 95, wherein the graphical representation of data is a histogram indicating the distribution of individuals across the range of clinical response values or other phenotype data.

99. The computer-usable medium of any one of claims 94,95, or 96, wherein at least one cell in the displayed matrix includes a selectable area, and wherein the medium further comprises computer-readable program code stored thereon for causing a computer to display, for individuals having the haplotype pair represented by the coordinates of the cell in the matrix, a histogram indicating the distribution of the individuals across the range of clinical response values.

100. The computer-usable medium of any one of claims 94,95, or 96, which further comprises computer-readable program code stored thereon for causing a computer to display a third selectable item, and computer-readable program code stored thereon for causing a computer to display, in response to selection of the third selectable item by the user, the statistical significance of the correlations between variation at individual polymorphic sites and the clinical response values.

101. The computer-usable medium of any one of claims 94,95, or 96, which further comprises computer-readable program code stored thereon for causing a computer to display a fourth selectable item, and computer-readable program code stored thereon for causing a computer to display, in response to selection of the fourth selectable item by the user, the numerical mean and standard deviation of clinical response values among individuals having each haplotype pair in the matrix.

102. The computer-usable medium of any one of claims 94,95, or 96, which further comprises computer-readable program code stored thereon for causing a computer to display a fifth selectable item, and computer-readable program code stored thereon for causing a computer to display, in response to selection of the fifth selectable item by the user, the results of an analysis of variation calculation to permit determination of whether variation in the clinical response values between individuals having different haplotype pairs is statistically significant.- 103. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to carry out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data for a gene or gene feature of interest to a clinical response measurement, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to display a variable controller for setting the number of genetic algorithm generations parameter; (b) computer-readable program code for causing a computer to display a variable controller for setting the number of agents parameter; (c) computer-readable program code for causing a computer to display a variable controller for setting the mutation rate parameter; (d) computer-readable program code for causing a computer to display a variable controller for setting the crossover rate parameter; (e) computer-readable program code for causing a computer to display one or more selectable items each corresponding to a polymorphic site of the gene or gene feature of interest; and (f) computer-readable program code for causing a computer to displaying a selectable item for initiation of the genetic algorithm calculation, and (g) computer-readable program code for causing a computer, in response to the selection by the user of one or more selectable items

corresponding to a polymorphic site, and selection by the user of the item for initiation of the genetic algorithm caclulation, to execute the genetic algorithm calculation with the parameters set by the variable controllers, and to display on a display device (i) the residual error of the model as a function of the number of genetic algorithm generations, and (ii) the results of the genetic algorithm calculation showing the optimal weights for each of the polymorphic sites.

104. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to display on a display device correlations between clinical outcome values obtained from selected clinical outome measures for a selected population, the computer-readable program code comprising: 6) (a) computer-readable program code for causing a computer to display a first plurality of selectable items corresponding to clinical outcome measurements; 7) (b) computer-readable program code for causing a computer to display a second plurality of selectable items corresponding to clinical outcome measurements; and 8) (c) computer-readable program code for causing a computer to display a scatter plot of data points, each data point corresponding to an individual in the selected population; 9) (d) computer-readable program code for causing a computer, in response to selection by the user of an item from among the first plurality of selectable items, to locate each data point along the x axis of the scatter plot according to the clinical outcome value for the associated individual from the clinical measurement represented by the selected item; and 10) (e) computer-readable program code for causing the computer, in response to selection by the user of an item from among the second plurality of selectable items, to locate each data point along the y axis of the scatter plot according to the clinical outcome value for the associated individual from the clinical measurement represented by the selected item.

105. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to provide information of use in conducting a clinical trial of a treatment protocol for a medical condition of interest, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database of DNA sequence data for selected genes or other loci in a reference population of individuals, and to access a database of (or accept as input) DNA sequence data for selected genes or other loci in a clinical trial population of individuals; (b) computer-readable program code for causing a computer to assign to each member of the reference population haplotypes for each of the selected genes or other loci; (c) computer-readable program code for causing a computer to calculate the frequencies, population distributions and statistical measures, including confidence limits, for each of the assigned haplotypes in the reference population; (d) computer-readable program code for causing a computer to assign to each member of a trial population haplotypes for each of the selected genes or other loci, based upon the frequencies, population distributions and statistical measures calculated in the reference population; (e) computer-readable program code for causing a computer to determinine the correlations between individual responses to the treatment and individual haplotypes, for each of the selected genes or other loci; (f) computer-readable program code for causing a computer to accept as input an individual's DNA sequence data or haplotypes for one or more of the selected genes or other loci; and (g) computer-readable program code for causing a computer to display or output the expected response of the individual to the treatment, based on the determined correlations between individual responses to the treatment and individual haplotypes.

106. The computer-usable medium of claim 105, which further comprises: (a) computer-readable program code stored thereon for causing a computer to derive from the haplotype distribution found for the reference population a reduced set of genotyping markers, which allow an individual's haplotypes to be accurately predicted without conducting a complete molecular haplotype analysis; and (b) computer-readable program code stored thereon for causing a computer to use the reduced set of genotype markers to assign haplotypes.

107. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to infer genotypes of individual subjects for a selected gene having at least m polymorphic sites, the computer-readable program codecomprising: (a) computer-readable program code for causing a computer to access a database of m-site haplotypes of the selected gene from a representative cohort of individuals; (b) computer-readable program code for causing a computer to tabulate the frequency of occurrence for each of the haplotypes; (c) computer-readable program code for causing a computer to construct a list of all genotypes that could result from all possible pairs of observed haplotypes; (d) computer-readable program code for causing a computer to calculate the expected frequency of these genotypes assuming the Hardy-Weinberg equilibrium; (e) computer-readable program code for causing a computer to generate a complete set of all possible masks of the same length m as the haplotypes, wherein each mask blocks the identity of the nucleotides at m-n polymorphic sites and admits the identity of nucleotides at the other n sites; (f) computer-readable program code for causing a computer to for calculate, for each mask, how much ambiguity results from genotyping with only the n polymorphic sites whose identity is admitted by the mask; (g) computer-readable program code for causing a computer to output or display on a display device the calculated ambiguity for one or more masks.

108. The computer-usable medium of claim 107, which further comprises computer-readable program code stored thereon for causing a computer to calculate the level of ambiguity for a mask, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to identify all pairs of genotypes that are rendered identical by application of the mask; (b) computer-readable program code for causing a computer to calculate the geometric mean of the calculated Hardy-Weinberg frequencies of each pair of genotypes rendered identical by application of the mask; (c) computer-readable program code for causing a computer to sum all such geometric means for all ambiguous pairs to obtain an ambiguity score for the mask.

109. The computer-usable medium of claims 107 or 108, which further comprises computer-readable program code stored thereon for causing a computer to assign a haplotype pair to an individual having an ambiguous genotype, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to calculate, for two haplotype pairs A and B that could explain a given genotype, the Hardy-Weinberg equilibrium probabilities PA and PB, where PA + PB = 1 ; (b) computer-readable program code for causing a computer to assign a haplotype pair by a process comprising (i) selecting a random number between 0 and 1; (ii) if the random number is less than or equal to PA, assigning the haplotype pair A; and (iii) if the number is greater than PA, assigning the haplotype pair B.

110. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database containing haplotype information, and clinical response or outcome data (clinical outcome values) or other phenotype data, from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each individual SNP in the haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and generating a numerical measure of the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those individual SNPs whose numerical measure of the degree of correlation with the clinical outcome values or other phenotype data exceeds a first cut-off value: (d) computer-readable program code for causing a computer to generate all possible pair-wise combinations of the saved SNPs so as to provide a set of n-site sub-haplotypes where n = 2; (e) computer-readable program code for causing a computer to statistically analyze each newly generated n-site sub-haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate a numerical measure of the degree of correlation; (f) computer-readable program code for causing a computer to store for further processing those n-site sub-haplotypes whose numerical measure of the degree of correlation exceeds the first cut-off value; (g) computer-readable program code for causing a computer to generate all possible pair-wise combinations among and between the saved SNPs and saved sub-haplotypes, to produce new subhaplotypes with increased values of n; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes can be generated, or (ii) no further sub-haplotypes having n less than a pre-selected or user-selected limit can be generated.

111. The computer-usable medium of claim 110, which further comprises computer-readable program code stored thereon for causing a computer to display those saved SNPs and sub-haplotypes whose numerical measure of the degree of correlation with the clinical outcome value or other phenotype exceeds a second cut- off value, wherein the second cut-off value is greater than the first cut-off value.

112. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database containing haplotype information, and clinical response or outcome data (clinical outcome values) or other phenotype data, from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each individual SNP in the haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate the p-value for the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those individual SNPs whose p-value for the degree of correlation does not exceed a first cut-off value; (d) computer-readable program code for causing a computer to generate all possible pair-wise combinations of the saved SNPs so as to provide a set of n-site sub-haplotypes where n = 2; (e) computer-readable program code for causing a computer to statistically analyze each newly generated n-site sub-haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate the p-value for the degree of correlation; (f) computer-readable program code for causing a computer to store for further processing those n-site sub-haplotypes whose p-value for the degree of correlation does not exceed the first cut-off value; (g) computer-readable program code for causing a computer to generate all possible pair-wise combinations among and between the saved SNPs and saved sub-haplotypes, to produce new subhaplotypes with increased values of n ; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-

haplotypes can be generated, or (ii) no further sub-haplotypes having n less than a pre-selected or user-selected limit can be generated.

113. The computer-usable medium of claim 110, which further comprises computer-readable program code stored thereon for causing a computer to display those saved SNPs and sub-haplotypes whose p-value for the degree of correlation with the clinical outcome value or other phenotype does not exceed a second cut-off value, wherein the second cut-off value is less than the first cut-off value.

114. The computer-usable medium of claims 110-113, which further comprises computer-readable program code stored thereon for causing a computer to exclude from further processing complex subhaplotypes which are constructed from smaller sub-haplotypes, where the smaller sub-haplotypes each have correlation values that are at least as significant as that of the complex sub- haplotype.

115. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype of interest, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database containing single gene haplotype information for one or more genes, and clinical response, outcome data, or other phenotype data from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each single gene haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and to generate a numerical measure of the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those haplotypes whose numerical measure of the degree of correlation exceeds a first cut-off value; (d) computer-readable program code for causing a computer to generate, for each haplotype composed of m polymorphic sites, all possible sub-haplotypes having a single site masked, so as to provide a set of m-n site sub-haplotypes where n = 1; (e) computer-readable program code for causing a computer to statistically analyze each newly generated sub-haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and calculating a numerical measure of the degree of correlation; (f) computer-readable program code for causing a computer to save for further processing those sub-haplotypes whose numerical measure of the degree of correlation exceeds the first cut-off value; (g) computer-readable program code for causing a computer to generate, from the saved sub-haplotypes, all possible sub- haplotypes having one additional site masked; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes have a degree of correlation which exceeds the first cut-off value, or (ii) no further sub-haplotypes having more unmasked sites than a pre-selected limit can be generated.

116. The computer-usable medium of claim 115, which further comprises computer-readable program code stored thereon for causing a computer to display those saved sub-haplotypes whose numerical measure of the degree of correlation with the clinical response data, outcome value, or other phenotype data exceeds a second cut-off value, wherein the second cut-off value is greater than the first cut- off value.

117. A computer-usable medium having computer-readable program code stored thereon, for causing a computer to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype of interest, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to access a database containing single gene haplotype information for one or more genes, and clinical response, outcome data, or other phenotype data from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each single gene haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and to calculate the p-value for the degree ofcorrelation; (c) computer-readable program code for causing a computer to store for further processing those haplotypes whose p-value for the degree of correlation does not exceed a first cut-off value; (d) computer-readable program code for causing a computer to generate, for each haplotype composed of m polymorphic sites, all possible sub-haplotypes having a single site masked, so as to provide a set of m-n site sub-haplotypes where n = 1; (e) computer-readable program code for causing a computer to statistically analyze each newly generated sub-haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and calculating the p-value for the degree of correlation; (f) computer-readable program code for causing a computer to save for further processing those sub-haplotypes whose p-value for the degree of correlation does not exceed the first cut-off value; (g) computer-readable program code for causing a computer to generate, from the saved sub-haplotypes, all possible sub- haplotypes having one additional site masked; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes have a p-value which does not the first cut-off value, or (ii) no further sub-haplotypes having more unmasked sites than a pre-selected limit can be generated.

118. The computer-usable medium of claim 117, which further comprises computer-readable program code stored thereon for causing a computer to display those saved sub-haplotypes whose p-value for the degree of correlation with the clinical response, outcome, or phenotype of interest does not exceed a second cut- off value, wherein the second cut-off value is less than the first cut-off value.

119. The computer-usable medium of claims 115-118, which further comprises computer-readable program code stored thereon for causing a computer to exclude from further processing complex sub-haplotypes which are constructed from smaller sub-haplotypes, where the smaller sub-haplotypes each have correlation values that are at least as significant as that of the complex sub- haplotype.

120. A computer programmed to cause haplotype pair assignments to be made to an individual member of a population whose genotype information for a gene or gene feature of interest is stored in a computer-readable form, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: computer-readable program code for causing a computer to generate all possible haplotype pairs consistent with the stored genotype; computer-readable program code for causing a computer to calculate the frequency of the haplotypes and haplotype pairs according to the Hardy-Weinberg equilibrium, based upon the observed distribution of haplotypes or haplotype pairs in the population; and computer-readable program code for causing a computer to select the most probable haplotype pair for the individual.

121. The computer of claim 120, wherein the program code further includes computer-readable program code for causing a computer to correct the stored distribution of haplotypes or haplotype pairs for effects imposed by the presence of a limited number of individuals in the population.

122. The computer of claim 120, wherein the program code further includes computer-readable program code for causing a computer to validate haplotype pair assignments by analyzing for compliance of the assigned haplotype pair with Mendelian inheritance principles.

123. The computer of claim 120, wherein the population is selected from the group consisting of a reference population, a clinical population, a disease population, an ethnic population, a family population and a same-sex population.

124. A computer programmed to cause haplotype pair assignments to be made to an individual member of a population whose genotype information for a gene or gene feature of interest is stored in a computer-readable form, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: computer-readable program code for causing a computer to generate all possible haplotype pairs consistent with the stored genotype; computer-readable program code for causing a computer to access a database containing reference haplotype pair frequency data and to determine from the frequency data the probability, for each of the possible haplotype pairs, that the individual has the possible haplotype pair; and computer-readable program code for causing a computer to select the most probable haplotype pair for the individual.

125. A computer programmed to identify a correlation between a clinical response to a treatment or other phenotype and a haplotype or haplotype pair present at a candidate locus hypothesized to be associated with the clinical response other phenotype, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database containing data on clinical responses to treatments, or other phenotypes, exhibited by individuals in a clinical population; (b) computer-readable program code for causing a computer to access a database containing haplotype data for each individual of the clinical population, the haplotype data comprising information on a plurality of polymorphic sites present at the candidate locus; and (c) computer-readable program code for causing a computer to calculate the degree of correlation between haplotypes or haplotype pairs and the clinical response to the treatment or other phenotype, by statistical analysis of the haplotype and clinical response data.

126. The computer of claim 125, wherein the treatment comprises administration of a drug or drug candidate.

127. The computer of claim 125, wherein the candidate locus is a gene or a gene feature.

128. The computer of claim 125, wherein the program code further includes computer-readable program code for causing a computer to store, display, or output the degree of correlation.

129. The computer of claim 125, wherein the program code further includes computer-readable program code for causing a computer to calculate the statistical significance of the correlation.

130. A computer programmed to identify a correlation between an individual's susceptibility to a condition or disease of interest, or other phenotype, and a haplotype or haplotype pair present at a candidate locus hypothesized to be associated with susceptibility to the condition or disease of interest, or with a phenotype of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access haplotype data for the candidate locus for each member of a population having the phenotype or condition or disease of interest ("disease haplotype data") ; (b) computer-readable program code for causing a computer to statistically analyze the disease haplotype data to calculate haplotype or haplotype pair frequencies; (c) computer-readable program code for causing a computer to access a database containing haplotype data for the candidate locus for each member of a healthy reference population ("reference haplotype data"); (d) computer-readable program code for causing a computer to statistically analyze the reference haplotype data to calculate haplotype or haplotype pair frequencies; and (e) computer-readable program code for causing a computer to identify a correlation of a haplotype or haplotype pair with susceptibility to the disease or condition of interest, or with the phenotype of interest, when the haplotype or haplotype pair has a higher frequency in the population having the phenotype, condition or disease of interest than in the reference population.

131. The computer of claim 130, wherein the candidate locus is a gene or a gene feature.

132. The computer of claim 130, wherein the program code further includes computer-readable program code for causing a computer to store, display, or output the identified correlation.

133. The computer of claim 130, wherein the program code further includes computer-readable program code for causing a computer to calculate the statistical significance of the correlation.

134. A computer programmed to predict an individual's response to a medical or pharmaceutical treatment based on one or more selected haplotypes or haplotype pairs of the individual, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database of correlations between haplotypes or haplotype pairs and responses to the medical or pharmaceutical treatment in a reference population; (b) computer-readable program code for causing a computer to locate haplotypes or haplotype pairs in the database that match the selected haplotypes or haplotype pairs of the individual, and (c) computer-readable program code for causing a computer to predict that the individual's response will be the response or responses associated in the database with the selected haplotype or haplotype pair.

135. The computer of claim 134, wherein the program code further includes computer-readable program code for causing a computer to generate an error estimate for the prediction.

136. A computer programmed to display a gene's structure and gene features on a display device, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to retrieve from a database, and display in a first area of the display device, data indicative of the frequencies of occurrence of a gene's haplotypes within predetermined member groupings of a reference population; (b) computer-readable program code for causing a computer to retrieve from a database data indicative of the gene's structure and gene features; (c) computer-readable program code for causing a computer to display in a second area of the display device a graphical representation of the gene's structure, user-selectable items indicating the location of gene features, and graphical indicators of the location of polymorphic sites on the gene; (d) computer-readable program code for causing a computer to display in a third area of the display device, in response to a user's selection of an item indicating a gene feature, a graphical representation of the structure of the gene feature having user- selectable items indicating the position of polymorphic sites; and (e) computer-readable program code for causing a computer to retrieve from a database, and display in a third area of the display device, in response to a user's selection of an item indicating

the position of a polymorphic site, data indicative of the frequencies within the member groupings of the occurrence of particular nucleotides at the polymorphic site.

137. A computer programmed to display on a display device haplotype pair frequency data within a population of individuals, for a selected gene or gene feature, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display on the display device a plurality of selectable items, each item corresponding to a polymorphic site in the gene or gene feature; (c) computer-readable program code for causing a computer to retrieve from a database and display on the display device, in response to a user's selection of one or more items indicating polymorphic sites, individual haplotype pairs in the database that differ at one or more of the selected polymorphic sites; and (d) computer-readable program code for causing a computer to display on the display device data indicative of the frequencies of the displayed haplotype pairs within one or more member groupings within the population.

138. A computer programmed to display on a display device polymorphic site linkage data for a gene or gene structure of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display on the display device one or more matrix structures, wherein the axes of each matrix structure represent the polymorphic sites in the gene or gene feature of interest, and wherein each matrix structure corresponds to a different population or population group; and (b) computer-readable program code for causing a computer to display on the display device, in each cell of a matrix structure, a graphical indication of degree of linkage between the twp polymorphic sites corresponding to the coordinates of the cell in the matrix.

139. The computer of claim 138, wherein color is used as the graphical indication of degree of linkage, and wherein the medium further comprises computer-readable program code for causing a computer to display a reference color scale relating color to degree of linkage.

140. A computer programmed to display on a display device a phylogenetic tree, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display a plurality of selectable items, each corresponding to a polymorphic site in the gene or gene feature of interest; and (b) computer-readable program code for causing a computer to display a phylogenetic tree structure having a node for each haplotype in a population, where the distance between nodes is proportional to the minimum number of nucleotides that would have to be changed to interconvert the corresponding haplotypes.

141. The computer of claim 140, wherein the program code further includes computer-readable program code for causing a computer to display connections between the nodes that indicate a single nucleotide difference between the haplotypes repesented by the nodes.

142. The computer of claim 140, wherein the program code further includes computer-readable program code for causing a computer to display at each node an indication of the relative frequency of occurrence of the haplotype represented by the node among different population groups.

143. A computer programmed to display a genotype analysis screen on a display device, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display a first plurality of selectable items, each corresponding to a polymorphic site, and a second plurality of selectable items, each corresponding to a polymorphic site; (b) computer-readable program code for causing a computer to display on the display device a matrix structure, wherein the axes of the matrix structure represent haplotypes in the gene or gene feature of interest that vary at the polymorphic sites selected from the first plurality of selectable items; and (c) computer-readable program code for causing a computer to display on the display device, in each cell of the matrix structure, a graphical indication of the reliability of the assignment to an individual of the haplotype pair corresponding to the coordinates of the cell in the matrix, when the individual is genotyped only at the polymorphic sites selected from the second plurality of selectable items.

144. The computer of claim 143, wherein color is used as the graphical indication of reliability of haplotype pair

assignment, and wherein wherein the program code further includes computer-readable program code for causing a computer to display a reference color scale relating color to reliability of haplotype pair assignment.

145. A computer programmed to display clinical response values, or other phenotype data, of a subject population as a function of haplotype pairs of the individuals in the population, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to retrieve from a computer-readable storage device, data representing haplotype pairs and clinical response values, or other phenotype data, for the subject population; and (b) computer-readable program code for causing a computer to graphically display a haplotype pair matrix structure, each of whose cells contains a graphical representation of the clinical response values or other phenotype data of individuals having the haplotype pair corresponding to the coordinates of that cell in the haplotype pair matrix.

146. A computer programmed to display on a display device clinical response values, or other phnotypic data, of a subject population as a function of the haplotype pairs of the individuals in the population for a gene or gene feature of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display one or more first selectable items representing polymorphic sites of the gene of gene feature; (b) computer-readable program code for causing a computer to display one or more second selectable items representing clinical measurements or phenotypes; and (c) computer-readable program code for causing a computer to display on the display device, in response to the selection by the user of at least one first and second selectable items, a haplotype pair matrix structure, wherein the axes of the matrix structure represent haplotypes in the gene or gene feature of interest that vary at the polymorphic sites corresponding to the first selected item or items, and wherein each of the cells of the matrix contains a graphical representation of the mean clinical response value, or other phenotype data, for the clinical measurement represented by the selected second item, of individuals having the haplotype pair corresponding to the coordinates of the cell in the haplotype pair matrix.

147. The computer of claim 145 or 146, wherein color is used as the graphical indication of mean clinical response value, or other phenotype data, and wherein the program code further includes computer-readable program code for causing a computer to display a reference color scale relating color to mean clinical response value.

148. The computer of claim 147, wherein the program code further includes: (a) computer-readable program code for causing a computer to display a means for adjusting the range of mean clinical response values or other phenotype data represented by the reference color scale; and (b) computer-readable program code for causing a computer, in response to the adjustment of the range of clinical response values or other phenotype data represented by the reference color scale, to adjust the color of the cells of the haplotype pair matrix.

149. The computer of claim 145 or 146, wherein the graphical representation of data is a histogram indicating the distribution of individuals across the range of clinical response values or other phenotype data.

150. The computer of any one of claims 145,146, or 147, wherein at least one cell in the displayed matrix includes a selectable area, and wherein the program code further includes computer-readable program code for causing a computer to display, for individuals having the haplotype pair represented by the coordinates of the cell in the matrix, a histogram indicating the distribution of the individuals across the range of clinical response values.

151. The computer of any one of claims 145,146, or 147 wherein the program code further includes computer-readable program code for causing a computer to display a third selectable item, and computer-readable program code for causing a computer to display, in response to selection of the third selectable item by the user, the statistical significance of the correlations between variation at individual polymorphic sites and the clinical response values.

152. The computer of any one of claims 145,146, or 147, wherein the program code further includes computer-readable program code for causing a computer to display a fourth selectable item, and computer-readable program code for causing a computer to display, in response to selection of the fourth selectable item by the user, the numerical mean and standard deviation of clinical response values among individuals having each haplotype pair in the matrix.

153. The computer of any one of claims 145,146, or 147, wherein the program code further includes computer-readable program code for causing a computer to display a fifth selectable item, and computer-readable program code for causing a computer to display, in response to selection of the fifth selectable item by the user, the results of an analysis of variation calculation to permit determination of whether variation in the clinical response values between individuals having different haplotype pairs is statistically significant.

154. A computer programmed to carry out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data for a gene or gene feature of interest to a clinical response measurement, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to display a variable controller for setting the number of genetic algorithm generations parameter; (b) computer-readable program code for causing a computer to display a variable controller for setting the number of agents parameter; (c) computer-readable program code for causing a computer to display a variable controller for setting the mutation rate parameter; (d) computer-readable program code for causing a computer to display a variable controller for setting the crossover rate parameter; (e) computer-readable program code for causing a computer to display one or more selectable items each corresponding to a polymorphic site of the gene or gene feature of interest; and (f) computer-readable program code for causing a computer to displaying a selectable item for initiation of the genetic algorithm calculation; and (g) computer-readable program code for causing a computer, in response to the selection by the user of one or more selectable items corresponding to a polymorphic site, and selection by the user of the item for initiation of the genetic algorithm caclulation, to execute the genetic algorithm calculation with the parameters set by the variable controllers, and to display on a display device (i) the residual error of the model as a function of the number of genetic algorithm generations, and (ii) the results of the genetic algorithm calculation showing the optimal weights for each of the polymorphic sites.

155. A computer programmed to display on a display device correlations between clinical outcome values obtained from selected clinical outome measures for a selected population, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: 11) (a) computer-readable program code for causing a computer to display a first plurality of selectable items corresponding to clinical outcome measurements; 12) (b) computer-readable program code for causing a computer to display a second plurality of selectable items corresponding to clinical outcome measurements; and 13) (c) computer-readable program code for causing a computer to display a scatter plot of data points, each data point corresponding to an individual in the selected population; 14) (d) computer-readable program code for causing a computer, in response to selection by the user of an item from among the first plurality of selectable items, to locate each data point along the x axis of the scatter plot according to the clinical outcome value for the associated individual from the clinical measurement represented by the selected item; and 15) (e) computer-readable program code for causing the computer, in response to selection by the user of an item from among the second plurality of selectable items, to locate each data point along the y axis of the scatter plot according to the clinical outcome value for the associated individual from the clinical measurement represented by the selected item.

156. A computer programmed to provide information of use in conducting a clinical trial of a treatment protocol for a medical condition of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database of DNA sequence data for selected genes or other loci in a reference population of individuals, and to access a database of (or accept as input) DNA sequence data for selected genes or other loci in a clinical trial population of individuals; (b) computer-readable program code for causing a computer to assign to each member of the reference population haplotypes for each of the selected genes or other loci; (c) computer-readable program code for causing a computer to calculate the frequencies, population distributions and statistical measures, including confidence limits, for each of the assigned haplotypes in the reference population; (d) computer-readable program code for causing a computer to assign to each member of a trial population haplotypes for each of the selected genes or other loci, based upon the frequencies, population distributions and statistical measures calculated in the reference population; (e) computer-readable program code for causing a computer to determinine the correlations between individual responses to the treatment and individual haplotypes, for each of the selected genes or other loci; (f) computer-readable program code for causing a computer to accept as input an individual's DNA sequence data or haplotypes for one or more of the selected genes or other loci; and (g) computer-readable program code for causing a computer to display or output the expected response of the individual to the treatment, based on the determined correlations between individual responses to the treatment and individual haplotypes.

157. The computer of claim 156, wherein the program code further includes: (a) computer-readable program code for

causing a computer to derive from the haplotype distribution found for the reference population a reduced set of genotyping markers, which allow an individual's haplotypes to be accurately predicted without conducting a complete molecular haplotype analysis; and (b) computer-readable program code for causing a computer to use the reduced set of genotype markers to assign haplotypes.

158. A computer programmed to infer genotypes of individual subjects for a selected gene having at least m polymorphic sites, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database of m-site haplotypes of the selected gene from a representative cohort of individuals; (b) computer-readable program code for causing a computer to tabulate the frequency of occurrence for each of the haplotypes; (c) computer-readable program code for causing a computer to construct a list of all genotypes that could result from all possible pairs of observed haplotypes; (d) computer-readable program code for causing a computer to calculate the expected frequency of these genotypes assuming the Hardy-Weinberg equilibrium; (e) computer-readable program code for causing a computer to generate a complete set of all possible masks of the same length m as the haplotypes, wherein each mask blocks the identity of the nucleotides at m-n polymorphic sites and admits the identity of nucleotides at the other n sites; (f) computer-readable program code for causing a computer to for calculate, for each mask, how much ambiguity results from genotyping with only the n polymorphic sites whose identity is admitted by the mask; (g) computer-readable program code for causing a computer to output or display on a display device the calculated ambiguity for one or more masks.

159. The computer of claim 158, wherein the program code further includes computer-readable program code for causing a computer to calculate the level of ambiguity for a mask, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to identify all pairs of genotypes that are rendered identical by application of the mask; (b) computer-readable program code for causing a computer to calculate the geometric mean of the calculated Hardy-Weinberg frequencies of each pair of genotypes rendered identical by application of the mask; (c) computer-readable program code for causing a computer to sum all such geometric means for all ambiguous pairs to obtain an ambiguity score for the mask.

160. The computer of any one of claims 158 or 159, wherein the program code further includes computer-readable program code for causing a computer to assign a haplotype pair to an individual having an ambiguous genotype, the computer-readable program code comprising: (a) computer-readable program code for causing a computer to calculate, for two haplotype pairs A and B that could explain a given genotype, the Hardy-Weinberg equilibrium probabilities PA and PB, where PA + PB = 1; (b) computer-readable program code for causing a computer to assign a haplotype pair by a process comprising (i) selecting a random number between 0 and 1; (ii) if the random number is less than or equal to PA, assigning the haplotype pair A; and (iii) if the number is greater than PA, assigning the haplotype pair B.

161. A computer programmed to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database containing haplotype information, and clinical response or outcome data (clinical outcome values) or other phenotype data, from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each individual SNP in the haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and generating a numerical measure of the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those individual SNPs whose numerical measure of the degree of correlation with the clinical outcome values or other phenotype data exceeds a first cut-off value; (d) computer-readable program code for causing a computer to generate all possible pair-wise combinations of the saved SNPs so as to provide a set of n-site sub-haplotypes where n = 2; (e) computer-readable program code for causing a computer to statistically analyze each newly generated n-site sub-haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate a numerical measure of the degree of correlation; (f) computer-readable program code for causing a computer to store for further processing those n-site sub-haplotypes whose numerical measure of the degree of correlation exceeds the first cut-off value; (g) computer-readable program code for causing a computer to generate all possible pair-wise combinations among and between the saved SNPs and saved sub-haplotypes, to produce new subhaplotypes with increased values of n ; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes can be generated, or (ii) no further sub-haplotypes having n less than a pre-selected or user-selected limit can be generated.

162. The computer of claim 161, wherein the program code further includes computer-readable program code for causing a computer to display those saved SNPs and sub-haplotypes whose numerical measure of the degree of correlation with the clinical outcome value or other phenotype exceeds a second cut- off value, wherein the second cut-off value is greater than the first cut-off value.

163. A computer programmed to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database containing haplotype information, and clinical response or outcome data (clinical outcome values) or other phenotype data, from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each individual SNP in the haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate the p-value for the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those individual SNPs whose p-value for the degree of correlation does not exceed a first cut-off value; (d) computer-readable program code for causing a computer to generate all possible pair-wise combinations of the saved SNPs so as to provide a set of n-site sub-haplotypes where n = 2; (e) computer-readable program code for causing a computer to statistically analyze each newly generated n-site sub-haplotype for the degree to which it correlates with the clinical outcome values or other phenotype data, and calculate the p-value for the degree of correlation; (f) computer-readable program code for causing a computer to store for further processing those n-site sub-haplotypes whose p-value for the degree of correlation does not exceed the first cut-off value; (g) computer-readable program code for causing a computer to generate all possible pair-wise combinations among and between the saved SNPs and saved sub-haplotypes, to produce new subhaplotypes with increased values of n; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes can be generated, or (ii) no further sub-haplotypes having n less than a pre-selected or user-selected limit can be generated.

164. The computer of claim 161, wherein the program code further includes computer-readable program code for causing a computer to display those saved SNPs and sub-haplotypes whose p-value for the degree of correlation with the clinical outcome value or other phenotype does not exceed a second cut-off value, wherein the second cut-off value is less than the first cut-off value.

165. The computer of any one of claims 161-164, wherein the program code further includes computer-readable program code for causing a computer to exclude from further processing complex subhaplotypes which are constructed from smaller sub-haplotypes, where the smaller sub-haplotypes each have correlation values that are at least as significant as that of the complex sub-haplotype.

166. A computer programmed to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database containing single gene haplotype information for one or more genes, and clinical response, outcome data, or other phenotype data from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each single gene haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and to generate a numerical measure of the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those haplotypes whose numerical measure of the degree of correlation exceeds a first cut-off value; (d) computer-readable program code for causing a computer to generate, for each haplotype composed of m polymorphic sites, all possible sub-haplotypes having a single site masked, so as to provide a set of m-n site sub-haplotypes where n = 1; (e) computer-readable program code for causing a computer to statistically analyze each newly generated sub-haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and calculating a numerical measure of the degree of correlation; (f) computer-readable program code for causing a computer to save for further processing those sub-haplotypes whose numerical measure of the degree of correlation exceeds the first cut-off value; (g) computer-readable program code for causing a computer to generate, from the saved sub-haplotypes, all possible sub- haplotypes having one additional site masked; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes have a degree of correlation which exceeds the first cut-off value, or (ii) no further sub-haplotypes having more unmasked sites than a pre-selected limit can be generated.

167. The computer of claim 166, wherein the program code further includes computer-readable program code for causing

a computer to display those saved sub-haplotypes whose numerical measure of the degree of correlation with the clinical response data, outcome value, or other phenotype data exceeds a second cut- off value, wherein the second cut-off value is greater than the first cut-off value.

168. A computer programmed to determine polymorphic sites or sub- haplotypes that correlate with a clinical response or outcome of interest, or other phenotype of interest, the computer comprising a memory having at least one region for storing computer executable program code and a processor for executing the program code stored in memory, wherein the program code includes: (a) computer-readable program code for causing a computer to access a database containing single gene haplotype information for one or more genes, and clinical response, outcome data, or other phenotype data from a cohort of subjects; (b) computer-readable program code for causing a computer to statistically analyze each single gene haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and to calculate the p-value for the degree of correlation; (c) computer-readable program code for causing a computer to store for further processing those haplotypes whose p-value for the degree of correlation does not exceed a first cut-off value; (d) computer-readable program code for causing a computer to generate, for each haplotype composed of m polymorphic sites, all possible sub-haplotypes having a single site masked, so as to provide a set of m-n site sub-haplotypes where n = 1; (e) computer-readable program code for causing a computer to statistically analyze each newly generated sub-haplotype for the degree to which it correlates with the clinical response, outcome, or phenotype of interest, and calculating the p-value for the degree of correlation; (f) computer-readable program code for causing a computer to save for further processing those sub-haplotypes whose p-value for the degree of correlation does not exceed the first cut-off value; (g) computer-readable program code for causing a computer to generate, from the saved sub-haplotypes, all possible sub- haplotypes having one additional site masked; (h) computer-readable program code for causing a computer to repeat steps (e) through (g) until either (i) no new sub-haplotypes have a p-value which does not the first cut-off value, or (ii) no further sub-haplotypes having more unmasked sites than a pre-selected limit can be generated.

169. The computer of claim 168, wherein the program code further includes computer-readable program code for causing a computer to display those saved sub-haplotypes whose p-value for the degree of correlation with the clinical response, outcome, or phenotype of interest does not exceed a second cut-off value, wherein the second cut-off value is less than the first cut-off value.

170. The computer of any one of claims 166-169, wherein the program code further includes computer-readable program code for causing a computer to exclude from further processing complex sub-haplotypes which are constructed from smaller sub-haplotypes, where the smaller sub-haplotypes each have correlation values that are at least as significant as that of the complex sub- haplotype.

171. A data structure for storing and organizing biological information, stored on a computer-readable medium and accessible by a processor, which comprises a single parent table which is adapted for storing, organizing, and retrieving a plurality of genetic features by the relative positional relationships between the genetic features.

172. The data structure of claim 171, wherein said parent table is part of each of three submodels comprising the data structure, wherein said submodels are a genomic repository submodel, a variation repository submodel and a literature repository submodel.

173. The data structure of claim 172, wherein the genetic features are selected from the group consisting of chromosomes, genomic regions, genes, gene regions, gene transcripts, transcript regions, and polymorphisms.

174. The data structure of claim 173, further comprising a clinical repository submodel.

175. The data structure of claim 174, further comprising a drug repository submodel.

176. A method for storing and organizing biological information, which comprises (a) providing a data structure comprising a single parent table which is adapted for storing, organizing, and retrieving a plurality of genetic features by the relative positional relationships between the genetic features; and (b) positioning a first genetic feature onto a second genetic feature.

177. The method of claim 175, wherein said first genetic feature is an assembly and said second genetic feature is a gene.

178. The method of claim 177, further comprising positioning a third genetic feature onto said gene.

179. The method of claim 178, wherein said third genetic feature is a gene region and the method further comprises positioning onto said gene region a polymorphism.

180. The method of claim 179, further comprising providing a relationship between the polymorphism and at least one phenotype which is associated with the polymorphism.

181. The method of claim 177, further comprising positioning onto said gene a haplotype which comprises a plurality of polymorphisms.

182. The method of claim 178, further comprising providing a relationship between the haplotype and at least one phenotype which is associated with the haplotype.

183. A data structure for storing and organizing biological information, stored on a computer-readable medium and accessible by a processor, which comprises at least two different fields, one of which includes a plurality of genetic features, and the other of which includes relative positional relationships between the genetic features.

WORLD INTELLECTUAL PROPERTY ORGANIZATION

IP SERVICES

Home > IP Services > PATENTSCOPE® > Patent Search

Search result: 1 of 1

(WO/2001/001218) METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

Biblio. Data | Description | Claims | National Phase | Notices | Documents

International Application Status

Date	Title		
03.08.2009	International Application Status Report	view	download

Published International Application

Date ▼	Title		
07.06.2001	Later publication of international search report (A3 23/2001)	view	download
04.01.2001	Initial Publication without ISR (A2 01/2001)	view	download

Related Documents on file at the International Bureau (more information)

Date ▼	Title		
06.12.2002	International Preliminary Examination Report	view	download

Search result: 1 of 1

(WO/2001/001218) METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

Biblio. Data | Description | Claims | National Phase | Notices | Documents

Available information on National Phase entries (more information)

Office Code	National Entry Date	National Reference Number	Status
AU	28.11.2001	56386/00	
CA	22.11.2001	2369485	
EP	18.01.2002	2000941722	Published: 29.05.2002 Withdrawn: 04.01.2005
JP	25.12.2001	2001507164	
US	03.08.2001	09923235	
US	17.05.2002	10019415	
US	21.12.2001	10019242	
US	21.12.2001	10019342	